We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter 15

Chromatographic Retention Parameters as Molecular Descriptors for Lipophilicity in QSA(P)R Studies of Bile Acid

Mihalj Poša

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/48356

1. Introduction

Beside of separation and identification of chemical compounds, chromatography can be used to obtain molecular parameters that reflect their structural characteristics – molecular descriptors. Most often it is a parameter of hydrophobicity (lipophilicity) of molecules (ions) which is obtained by using retention parameters of reversed phase liquid chromatography or thin layer chromatography of high resolutions (RPHPLC and RPTLC) [1,2]. Hydrophobicity of molecules is an important feature in medical chemistry, and arbitrarily connected with the logarithm of the solute partition (distribution) coefficient log*P* (*P* stands for the ratio of the equilibrium concentration of the particle of the same electronic structure in 1-octanol and its equilibrium concentration in water). Partition coefficient is determined by the traditional shake flask method, which has drawbacks: long analysis (for reaching equilibrium) and the results often do not have adequate reproducibility [1,2,3]. While hydrophobicity obtained by chromatographic methods is obtained relatively quickly, it is possible to specify a number of compounds and achieved by high precision and reproducibility of results. Lipophilicity of chemical compounds (which is expressed either as a chromatographically i.e. retention parameters or as an *in silico* molecular descriptors of log*P*) is often included in the regression equations obtained in the QSA(P)R (Quantitative Structure Activity (Property) Relationship) studies [1,2,3].

Bile acids are amphiphilic molecules that have peculiar structure, because molecular descriptors that are obtained on the basis of the molecular graph or fragmentation methods often do not reflect their true structural features [4]. Therefore, the bile acid chromatographic lipophilicity play an important application in obtaining QSA(P)R models that connect biological and pharmacological or other physical-chemical properties

(solubility, critical micelle concentration, critical micelle temperature, etc.) for their structure [5]. In the QSA(P)R models chromatographic parameters are independent variables.

In the following section presents the introduction chromatographic parameters (TLC and HPLC) that are used to represent the hydrophobicity of compounds, and presents the main structural features of bile acids.

2. Chromatographic parameters

The basic characteristic of the position of analytes spot on TLC chromatograms is the R_f value (retardaction factor). R_f value is the ratio of the path length that has crossed spot of solute from the start line S_s and path length S_f of the solvent front (Fig. 1).

$$R_f = \frac{s_s}{s_f} \tag{1}$$

If the solute spends more time in the mobile phase, then its chromatographic spot position is closer to the front of solvent and its R_f value is even higher (maximum value of the R_f parameter is 1). Whereas if the solute spends more time in the stationary phase, then the position of his spots closer to the start line, and its R_f value is less (Fig. 1 (a)).

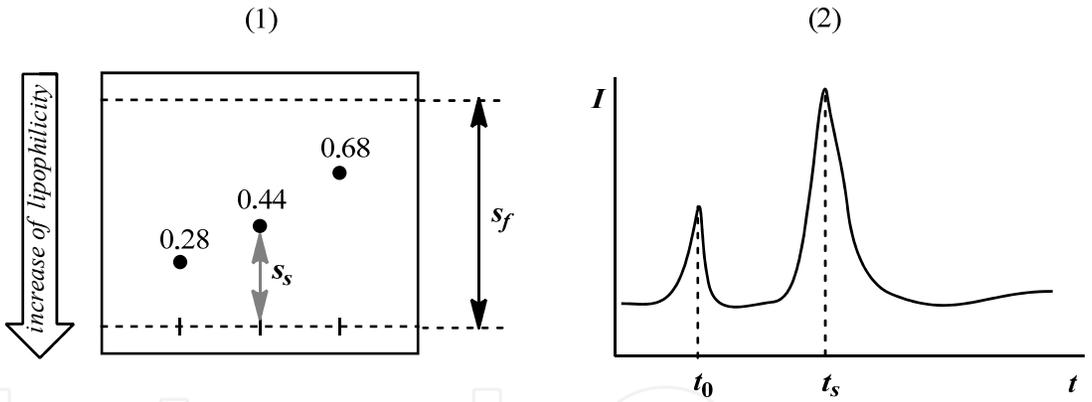


Figure 1. Chromatograms: (1) TLC with R_f values of the solute over their spots, (2) HPLC.

The ratio of time spent by the solute in the stationary phase t_s and the time by spend in the mobile phase t_m is the capacitance factor.

$$k = \frac{t_s}{t_m} \tag{2}$$

The connection between R_f values and capacitive factor has the equation:

$$k = \frac{1}{R_f} - 1 \tag{3}$$

The logarithm of the above expression is:

$$\log k = \log\left(\frac{1}{R_f} - 1\right) \tag{4}$$

and the retention parameter log*k* usually indicates with R_M. In the reverse phase thin layer chromatography, where the stationary phase is the hydrophobic environment, in the course of the chromatographic process chemical compounds that are more hydrophobic spend more time in the stationary phase than in the polar mobile phase, which resulting in less crossed paths of their chromatographic spots. Therefore, the more hydrophobic is solute its R_f value is more lower or its retention parameter R_M value is more higher. In reverse phase thin-layer chromatography for each compound is determined the dependence of the chromatographic parameters R_M of the volume fraction ϕ of organic modifier in aqueous mobile phase (i.e. for each of the tested compounds from one chromatographic experiment to another experiment varies of mobile phase volume fraction of organic modifier). The most common organic modifier is methanol. If increasing the volume fraction of the organic modifier in aqueous mobile phase results of decrease in hydrophobicity of mobile phase. This is manifested as the reduction of the difference between staying time of solute in the stationary and mobile phase during the chromatographic process – increased R_f values (decreasing value of R_M). Usually between the chromatographic parameters R_M and the volume fraction of organic modifier ϕ is linear relation (usually in the interval: $0.2 \le \phi \le 0.8$) [3,6-8].

$$R_M = R_{M0} + S\varphi \tag{5}$$

In the above equation R_{M0} is the extrapolated value of the chromatographic parameters R_M which is governed to the mobile phase contain only water (or buffered aqueous solution), i.e. mobile phase without an organic modifier, while *S* is the slope of the right (Fig. 2). *S* is directly related to the specific surface of the stationary phase.

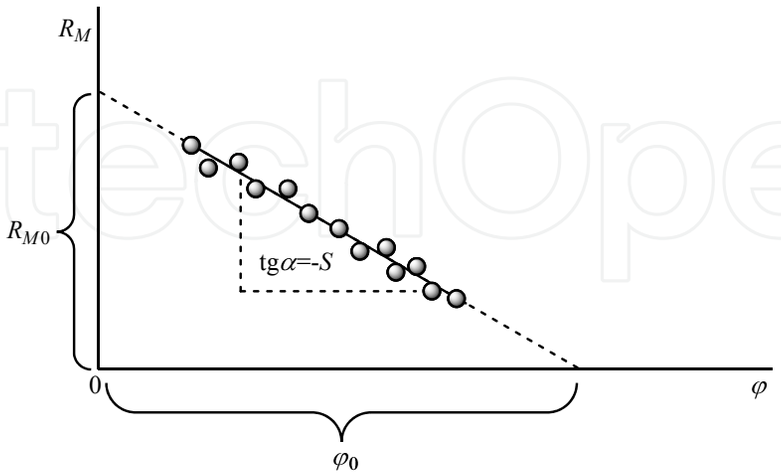


Figure 2. The linear dependence of R_M chromatographic parameters of volume of organic modifier ϕ.

R_{M0} parameters value depends on the type of organic modifier on the basis of which is determined by the function (1.5). R_{M0} is usually in the good correlation of the compounds

lipophilicity. In recent times to describe the molecules lipophilicity also used the chromatographic parameters ϕ_0 which is the ratio of R_{M0} and the slope S (Fig. 2) [7,8].

$$\varphi_0 = \frac{R_{M0}}{|-S|} \tag{6}$$

The chromatographic parameters ϕ_0 corresponding volume ratio of organic modifier in aqueous mobile phase in which the same amount of solute in the mobile phase and in the stationary phase. Indeed solute in the above mobile phase composition, during the chromatographic process, spends at the same time in the stationary phase and the mobile phase, therefore capacitive factor (2) is $k = 1$, respectively R_f value (3) is 0.5. This R_f values corresponding to R_M chromatographic parameter (4) whose value is zero. Which means that equation (5) is: $0 = R_{M0} + S\phi_0$ from which follows the expression (6). If the solute is more hydrophobic then a larger quantity of organic modifier is needed to equalize the amount of solute in two phases, i.e. the chromatographic parameter ϕ_0 has higher value. With a high efficient reverse phase liquid chromatography, stationary phase is also hydrophobic environment. Solute is characterized by retention time t_s, which represents the elapsed time from injection to the occurrence of the same solute in detectors, i.e. the retention time of solute in the column. Chemical compounds in RP-HPLC analysis is usually characterized by the retention coefficient (capacity factor):

$$k = \frac{t_s - t_0}{t_0} \tag{7}$$

where t_0 is the retention time of solvent from the mobile phase (Fig. 1 (a)). If a chemical compound is more hydrophobic, then more time is spent in the hydrophobic stationary phase, i.e. it takes more time to pass the column and the retention time or retention factor even greater. The logarithm of the retention coefficient log$k$ is used as a parameter of hydrophobicity of chemical compound. An important chromatographic parameter of lipophilicity is logk_w which were obtained an extrapolation of the linear equations (8) to the zero volume fraction ϕ of organic modifier in aqueous mobile phase.

$$\log k = \log k_w + S\varphi \tag{8}$$

Similar as in the RP TLC analysis, in RP HPLC also can defined the chromatographic indices ϕ_0 (with the same meaning) as the ratio of logk_w and the slope from log$k = f(\phi)$ [9,10].

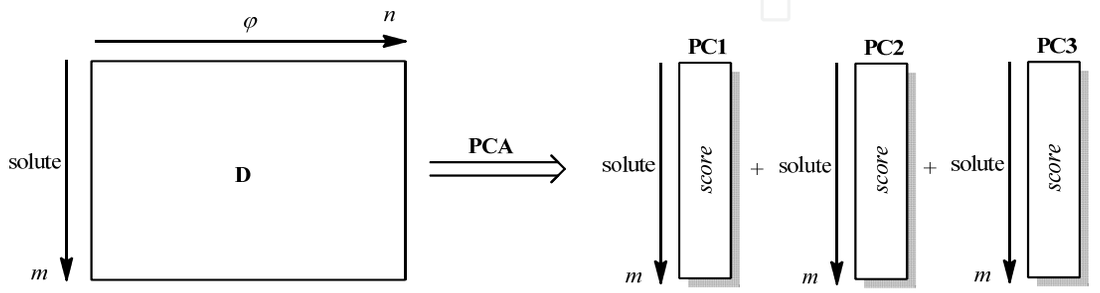


Figure 3. Application of principal component analysis on the retention data matrix D, n = number of different volume fraction of organic modifier, m = number of different compounds, n> 3.

There is a possibility of application principal component analysis (PCA) on chromatographic data to the thin layer and the liquid chromatography. PCA is applied to the data matrix **D** of retention parameters R_M, log$k$ (or k). The columns of the matrix **D** corresponding retention parameters for different volume fraction ϕ of organic modifier (columns represent the organic modifier) while the matrix rows represent the different types of chemical compounds. PCA is usually applied directly on the covariance matrix $\mathbf{D^TD}$ retention parameters. As a result of PCAs mathematical procedure (orthogonal diagonalization) are obtained orthonorms PC score vectors of whose number is 2 (PC1, PC2) or 3 (PC1, PC2, PC3), depending on the percentage of variance explained from the data matrix **D**. Accordingly objects (tested molecules) from a multidimensional space matrix of retention parameters **D** mapped to 2d or 3d space of PC. In the graphs of PC scores can be found congeneric group of chemical compounds by their lipophilicity. As a parameter of lipophilicity of chemical compounds in the QSA(P)R studies can be applied also to the scores of PC [3,6].

3. The hydrophobic effect

Generally, the dissolution of an amphiphilic object in water is accompanied by the disruption of the hydrogen bonds between water molecules and formation of a hydration sheath (hydration layer) around the particles of the dissolved substance. If we observe such a system (amphiphilic solution), which consists of two subsystems: hydrophobic part (**a**) and hydrophilic part (**b**), then the thermodynamic functions can be considered separately for each subsystem (Fig. 4). For both subsystems it holds that in the formation of the solvation sheaths (around the amphiphilic objects) approximately the same number of hydrogen bonds are formed as existed between the water molecules in the bulk water, i.e. before introducing the amphiphilic object. Hence the change of the enthalpy for each of the subsystems is equal to zero: $\Delta H(\mathbf{a}) \approx \Delta H(\mathbf{b}) \approx 0$. Also, in the formation of the solvation sheath, the entropy (translational and rotational) of both subsystems decreases, that is: $\Delta S(\mathbf{a}) \approx \Delta S(\mathbf{b}) < 0$. However, the water molecules from the hydrophilic side (**b**) of the amphiphilic molecule form additional hydrogen bonds; hence for this subsystem there is an additional negative enthalpy ($\Delta H_\varepsilon < 0$), which is dissipated as heath in the environment (bulk of the solution), thus giving rise to a positive change of the entropy (of the environment). Therefore, the free enthalpy change for the hydrophilic subsystem is: $\Delta G(\mathbf{b}) < 0$, on the basis of which the water molecules from the hydrophilic side (**b**) of the amphiphil can be denoted as stabilized water molecules (SWM), while the water molecules from the hydrophobic side (**a**) are nonstabilized water molecules (NSWM) [4,11].

The ratio of the hydrophilic-to-hydrophobic surface area of an amphiphilic molecule determines the overall change of the Gibbs energy of formation of the hydration sheath (ΔG), and, since the hydrophobic surface of the amphiphilic molecule is larger, then $\Delta G > 0$. The larger the amount of the amphiphil present in the solution, the more water molecules participate in the formation of the hydration sheath, and the more negative is the overall entropy change. This results in the changes in the system (solution) due to the passing of NSMW from the amphiphil hydration sheath to the bulk of the solution, giving rise to the

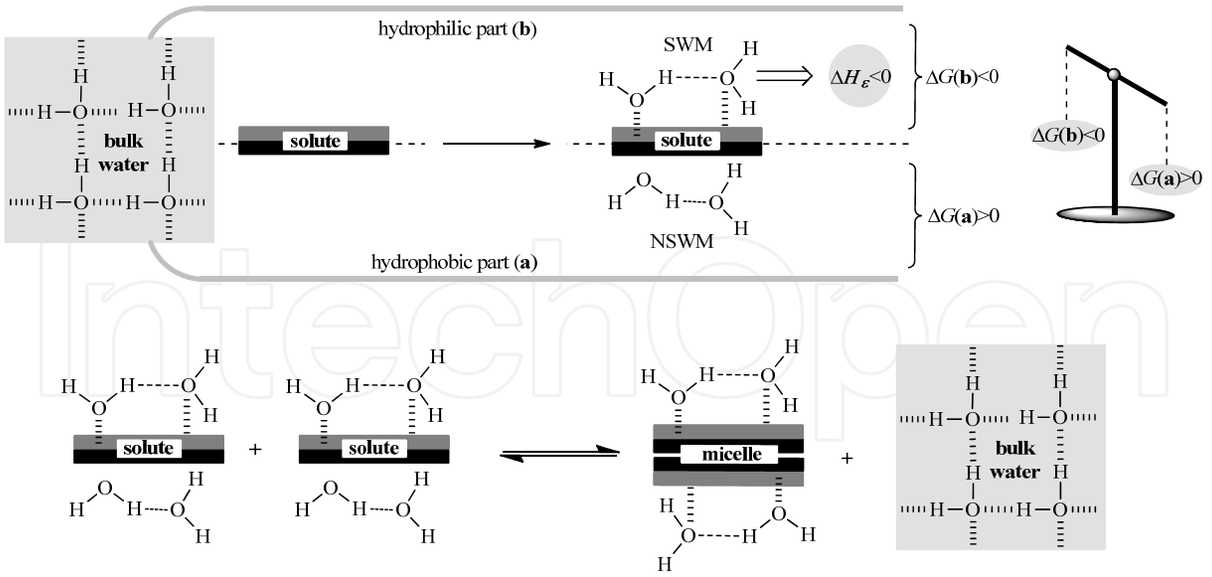


Figure 4. The hydrophobic effect: SWM = stabilised water molecules, NSWM = nonstabilised water molecules.

system entropy. Amphiphilic molecules (ions) are oriented so that their (formally) desolvated hydrophobic sides are to the smallest extent exposed to water molecules. This is realized either in the form of self-association of the amphiphils via their hydrophobic surfaces (Fig. 4) (entropy decrease due to the amphiphil self-association is smaller than the increase in the entropy due to their release from the hydration layer) or by binding of the amphiphil to the hydrophobic surface of the system: hydrophobic stationary phase, proteins, distribution in the organic solvent, etc. [11]. Generally, the hydrophobicity of a particle (molecule, ion) can be expressed via the logarithm of the coefficient of its partition (log*P*) between 1-octanol and water (*P* stands for the ratio of the equilibrium concentration of the particle of the same electronic structure in 1-octanol and its equilibrium concentration in water). If log*P* > 1, the particle is hydrophobic (lipophilic), and if log*P* < 1, the particle is hydrophilic [12]. Apart from the partition coefficient, the hydrophobicity of a particle can be also expressed via chromatographic parameters, either of the normal-phase or reversed-phase chromatography.

4. Structure of bile acids

Bile acid enzymatic produced in the liver of human and other mammals are the primary bile acids, and their intestinal microbial transformation to obtain the secondary bile acids. Both groups of bile acids are hydroxy derivatives of 5β-cholanic acid (cholic acid and chenodeoxycholic acid) [13]. The geometry of the 5β-cholanic acid largely determines of bile acids properties. The steroid skeleton of 5β-cholanic acid can be distinguished convex β and concave *α* surface (Fig. 5) [5].

Hydroxyl groups of bile acids are mostly oriented towards *α* side of steroid skeleton. X ray diffraction is demonstrated that the carboxyl group of the side chain from C17 carbon is also located on the *α* side of the steroid ring system. A concave (*α*) of bile acids steroid

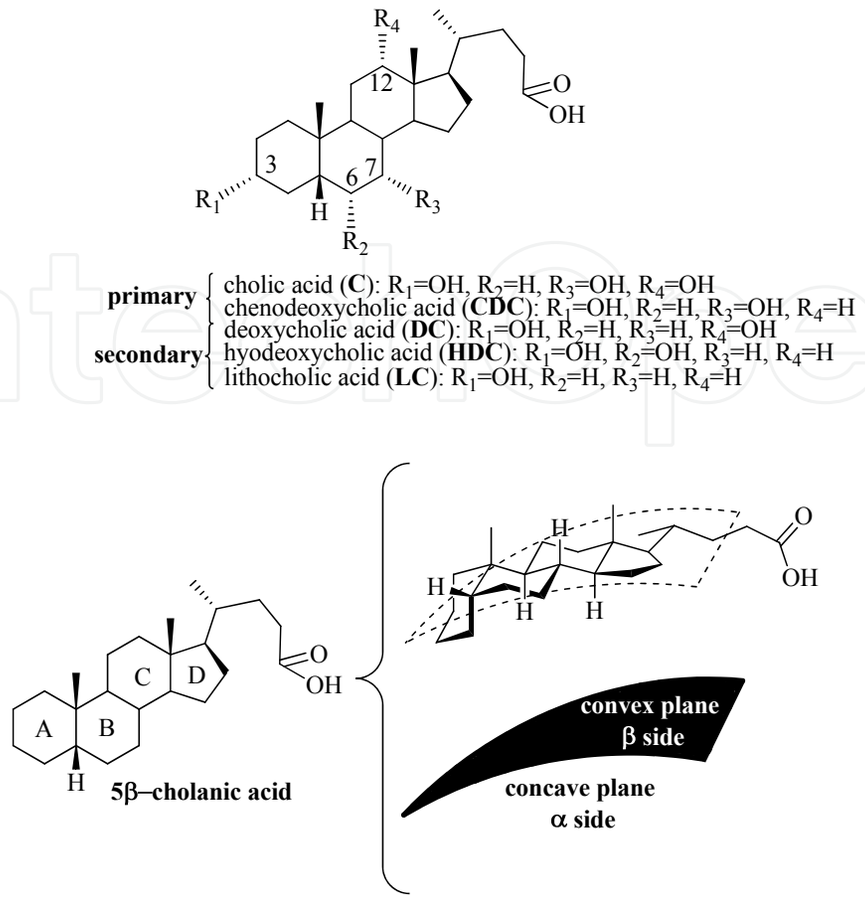


Figure 5. The bile acids general structure.

skeleton is the polar surface – hydrophilic side, while the convex surface (β) is the nonpolar surface – hydrophobic side. Simultaneous presence of hydrophobic and hydrophilic regions in the molecule of bile acids is named as amphiphilic. Since the carboxyl group ionized under physiological conditions, bile acids in biochemical systems belong to the ionic amphiphile. Also, bile acids are a special group of amphiphilic compounds, the group of biplanar molecules [4.5]. This is most apparent in cholic acid molecules (Fig. 6) because the molecules oxygen atoms from the α OH groups are in the same plane, the polar plane. In chenodeoxycholic and deoxycholic acid cannot be talking about the true planar polarity, but the literature often uses the term molecule with a hydrophilic edge.

If it oxidation of C7 or C12 α axial hydroxy group in cholic acid (**C**) (or deoxycolic acid (**DC**) or chenodeocycholic acid (**CDC**)) molecule leads to oxo derivative whose oxigen atom is switched for 60° (Newman projection) in relation to its' starting axial orientation i.e forms an angle of 30° with steroid skeleton mean plain (SSMP) (Newman projection, **Fig. 7**) [11,14]. Oxigen atom from C7 or C12 oxo group has the same steric orientation as has the equatorial C6 OH group from hyodeoxycholic acid (**HDA**).

Convergence of oxigen core (from oxo group) toward steroid sceleton β side means that there is the rise in number of stabilised water molecules (SWM, stabilised by hydrogen bonds) in hydratation layer of bile acids' steroid sceleton while hydrophobic surface of the

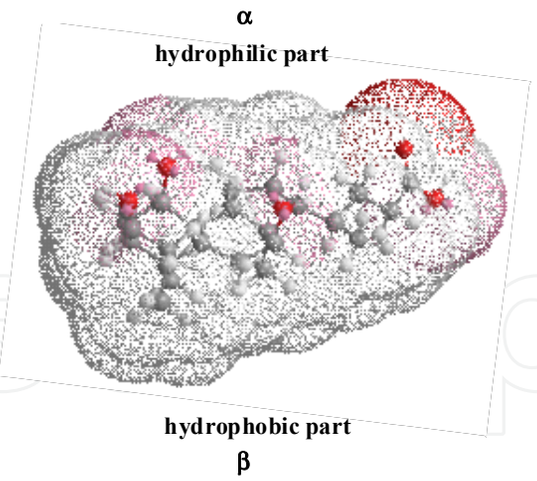


Figure 6. Two different side of cholic acid.

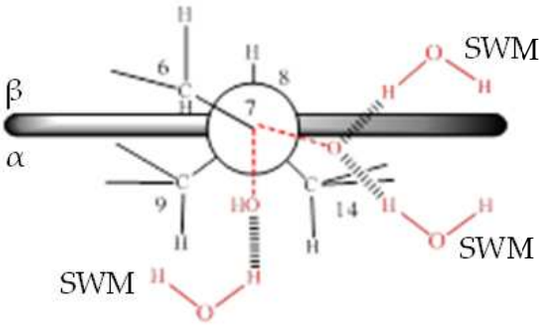


Figure 7. Introduction of oxo group in bile acids' steroid sceleton leads to rise in number of stabilised water molecules (SWM) i.e. the hydrophobic β side of the steroid sceleton decreases.

convex side of the steroid sceleton decreases (Fig. 7). Thus, the number of unstabilised water molecules (NSWM) declines [11].

$$\mathrm{BA}\left[\mathrm{NSWM}\right]_{\mathrm{n}(aq)} \rightleftarrows \mathrm{BA}_{(s)} + n\mathrm{NSWM} \qquad \text{(R1)}$$

If bile acids' hydrophobic surface deceases (B), i.e. the number of NSWM is lower, the observed molecule bonds to the hydrophobic stationary phase with lower intensity, i.e. reaction (R1) is moved to the right. This change in hydrophobic surface reflects in the values of retention coefficient k (RPHPLC) and retention parameter R_{M0} (RPTLC).

5. Discussion: bile acids and the hromatographic parametars (indices)

In the RPHPLC(TLC), there is an equilibrium between the bile acid molecules (ions) from the polar mobile phase $\mathrm{BA}_{(aq)}$ and the bile acid $\mathrm{BA}_{(s)}$ adsorbed on the hydrophobic stationary phase.

$$\mathrm{BA}_{(aq)} \rightleftarrows \mathrm{BA}_{(s)} \qquad \text{(R2)}$$

This equilibrium is characterized by the equilibrium constant K_{ad} = [$BA_{(s)}$] / [$BA_{(aq)}$], which is connected with the retention factor (capacity factor) (7) via the following relation [15]:

$$k = \frac{K_{ad} V_{hc}}{V_{pf}} = K_{ad} \varphi \tag{9}$$

where V_{hc} represents the volume of hydrocarbon in the stationary phase, whereas V_{pf} is the volume of the polar phase. Their ratio is denoted as ϕ, and it represents a characteristic of the column. The change of the standard Gibbs energy ΔG°_{ad} of the reaction (R2), taking also into account (9), is:

$$\Delta G^{\circ}_{ad} = -RT \ln K_{ad} = -RT \ln k + RT \ln \varphi \tag{10}$$

In both RPHPLC and RPTLC, in the adsorption of bile acids on the hydrophobic stationary phase, the water molecules from the solvation sheath that are not stabilized by hydrogen bonds (NSWM) pass to the hydrophilic mobile phase, giving a positive entropy contribution. Beside that, during the return of these water molecules in the bulk solution, NSWMs surface energy is released, which is then dissipated as a heat in the environment (solvophobic theory: according to this theory change in surface energy is a key factor in the reversed phase chromatographic processes energy). The formally dehydrated hydrophobic surfaces of bile acid molecules (ions) bind to the surface of the adsorbent by hydrophobic interactions (London dispersion forces, and dipol-induced dipole interactions). This makes a negative enthalpic contribution to the adsorption process, which is also dissipated as heat into the mobile phase (environment), giving rise to a positive change of the entropy of the environment.

Hence, the overall entropy change is positive, i.e. the Gibbs energy is lowered. The decrease in the free enthalpy of adsorption ($\Delta_{ad}G^{o}$) of bile acids (molecules or ions) on the hydrophobic stationary phase is more pronounced if the solvation layer of the steroid skeleton contains a greater number of NSWM, since then, there exists a largest increase in the overall entropy. In the binding of bile acids (molecules or ions) to the hydrophobic stationary phase, their steroid skeletons orient in such a way that the formally desolvated surfaces are to the smallest possible extent exposed to the solvent molecules from the hydrophilic mobile phase (methanol and water). Since bile acids are biplanar compounds, their molecules bind to the hydrophobic stationary phase by the side of the steroid skeleton that is most hydrophobic. With natural bile acids and their oxo derivatives, this is usually the convex β side of the steroid skeleton (an exception is the enantiomer of cholic acid with β configurations of all three OH groups, where α side is more hydrophobic than the β side of the steroid skeleton). Thus, the ln$k$ chromatographic parameter has the liner dependence of $\Delta_{ad}G^{o}$, which means that lnk has also linear depending on the hydrophobic surface of the β side of steroid nucleus. Similar is true for R_M chromatographic parameter (RPTLC). Therefore, ln$k$ could serve as a molecular descriptor for the hydrophobicity of bile acids. But the studies dealing with the hydrophobicity of bile acids as expressed via the retention coefficients (k), there is a problem of mutual comparability of the ln$k$ values. Because of that

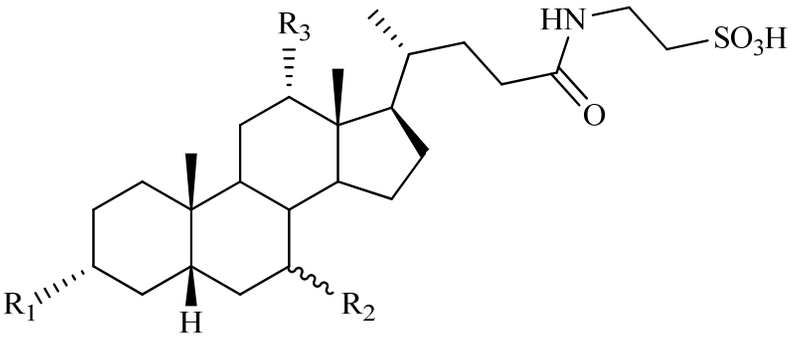
it is usually accustomed to give an order of hydrophobicity of the molecules, while the ln*k* values, though determined under identical experimental conditions, may vary because of the characteristic of the column (RPHPLC), which is in equation (9) denoted with ϕ. Because of that, Heuman introduced for bile acids a relative retention coefficient (k_r) which is defined as the ratio of the retention coefficient of the given bile acid (k(BA)) and the retention coefficient of taurocholic acid (**TC**) (k(TC)) (Fig. 8) [16].

$$k_r(\mathrm{BA}) = \frac{k(\mathrm{BA})}{k(\mathrm{TC})} \tag{11}$$

Taurocholic acid (**37**) was chosen as a reference because its retention coefficient in the pH range of the mobile phase from 2.9 to 9.0 is practically constant (complete ionization) [16]. By combining equations (10) and (11), the free enthalpy of adsorpion of bile acid (BA) on the hydrophobic stationary phase is:

$$\Delta G^{\circ}_{ad}(\mathrm{BA}) = -RT \ln k_r(\mathrm{BA}) + \Delta G^{\circ}_{ad}(\mathrm{TC}) \tag{12}$$

In this way are eliminated the individual characteristics of the column, so that the relative retention coefficient depends only on the composition of the mobile phase and the degree of hydrophobicity of the stationary phase.



taurocholic acid-**TC** (**37**): R_1=OH, R_2=α–OH, R_3=OH
taurolithocholic acid-**TL** (**38**): R_1=OH, R_2=H, R_3=H
tauroursocholic acid-**TU** (**39**): R_1=OH, R_2=β–OH, R_3=OH
tauroursodeoxycholic acid-**TUD** (**40**): R_1=OH, R_2=β–OH, R_3=H
taurodeoxycholic acid-**TD** (**41**): R_1=OH, R_2=H, R_3=OH

Figure 8. Tauro conjugated bile acids.

Besides, Heuman defined also the hydrophobicity index *HI* as the ratio of the logarithm of the relative retention coefficient of the given bile acid (BA) and of the relative retention coefficient of taurolithocholic acid (**TL**) (lnk_r(TL)) (Fig. 7):

$$HI(\mathrm{BA}) = \frac{\ln k_r(\mathrm{BA})}{\ln k_r(\mathrm{TL})} \tag{13}$$

By introducing equation (13) into (12), one obtains the following equation:

$$\Delta G^{\circ}_{ad}(\mathrm{BA}) = HI(\mathrm{BA}) \times \left(-RT \ln k_r(\mathrm{TL})\right) + \Delta G^{\circ}_{ad}(\mathrm{TC}) \tag{14}$$

from which is evident that there is a linear relation between the hydrophobicity index of a bile acid and the Gibbs energy of adsorption on the hydrophobic stationary phase.

The *HI* value for taurocholic acid (**TC**) is zero since its relative retention coefficient equation (11) is equal to one, and the same value (one) holds also for taurolithocholic acid (**TL**). If a certain bile acid is more hydrophilic than taurocholic acid (**TC**), its retention constant (*k*) is smaller than the *k* value of taurocholic acid (**TC**), so that the relative retention coefficient of that bile acid is smaller than unity (ln*k*<0), which means that, according to equation (13), the *HI*(BA) is negative. In the opposite case, if a certain bile acid is more hydrophobic than taurocholic acid (**TC**), then the value of its retention coefficient is greater than the retention coefficient of taurochollic acd (**37**), and the relative retention coefficient for that bile acid is larger than unity (ln*k*>0), that is *HI*>0 (Fig. 9).

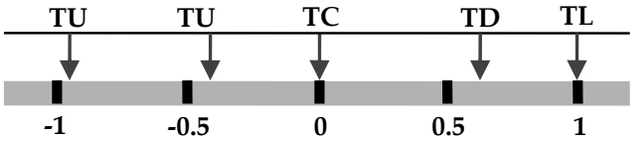


Figure 9. Heumans scale of bile acids hydrophobicity.

Besides of Heumans hydrophobic index (based on data retention) log*k* is used to obtain predictive equations where log*k* is the independent variable while bile acid properties associated with its hydrophobicity is the dependent variable. Armstrong and Carey were given a linear regression equation, where log*k* is associated with the solubilized molar fraction of cholesterol ϑ (solubilized by micellar solutions of bile acid salts in the equilibrium conditions) [17]:

$$\log k = S \log \vartheta - const. \tag{15}$$

In the plane of ϑ and log*k* is obtained of three linear congeneric groups of examined bile acids, a particular group form the non-conjugated bile acids, whereas the glyco- and tauro-conjugated bile acids are also in separate groups. For each of the derived linear congeneric group is determined equation (15) whose regression coefficients (R) is range in 0993-0999. Armstrong and Carey were predicted the capacity of cholesterol solubilization by bile acids oxo derivatives using the equations (15) for each congeneric group of bile acid [17]. It can be shown that log*k* is indeed associated with a number of hydration layers non-stabilized water molecules n_{NSWM} from β side of steroid skeleton:

$$\frac{d \ln k}{d\left(\Phi m_{NaCl}\right)} = n_{NSWM} M_{H_2O} \tag{16}$$

where Φm_{NaCl} is the product of the osmotic coefficient and the molality of NaCl, and M_{H_2O} is the molecular weight of water [11]. Therefore with the increasing of the number of the α

axial OH groups which is replacement with equatorial OHs or oxo groups resulting in a decrease of NSWM number from the hydratation shell of bile acids, which are then reflected in the increase of log*k* values i.e. increasing hydrophobicity of bile acids. However, the cholesterol solubilization depends on the size of hydrophobic domain of micelles, which depends on the hydrophobic surface of the β side of steroid nucleus, that's all mean that the log*k* adequately describes of cholesterol (or any other hydrophobic guest molecules) solubilization. However it should be noted that the use of log*k* as molecular descriptors, log*k* for each molecule, as the molecules from teaching set and as the molecules from the control set must be obtained in an identical column. Contrary to the different values of ϕ (9) increases the error of prediction.

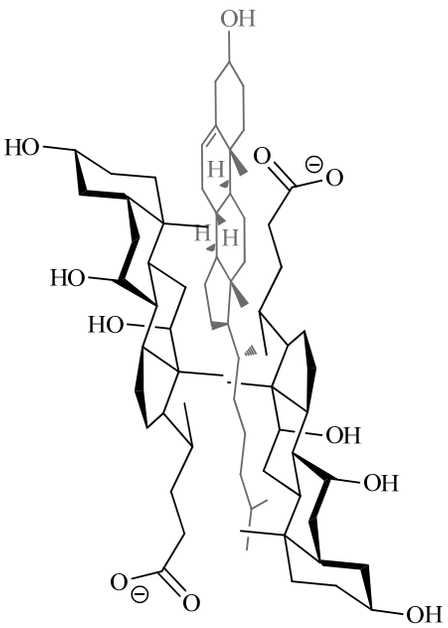


Figure 10. Micellar solubilization of cholesterol: cholesterol is in the contact with the hydrophobic surfaces of β side of bile acids steroid skeleton.

Natalini et al. as an experimental molecular descriptor used chromatographic parameter ϕ_0 (6) for modeling the negative logarithm of the critical micellar concentration (*pCMC* = -log*CMC*) [9,10]. The study was included 20 non-conjugated and conjugated bile acids. For the full set of bile acids were obtained following linear regression equation:

$$pCMC = 3.85\varphi_0 - 0.82 \tag{17}$$

$$R = 0.8246$$

Regression coefficient is improved if the equations of linear regression obtained separately for conjugated (18) and for non-conjugated bile acids (19):

$$pCMC = 4.78\varphi_0 - 1.57 \tag{18}$$

$$R = 0.9643$$

$$pCMC = 1.05\varphi_0 + 1.32 \quad (19)$$

$$R = 0.9380$$

Equations (18) and (19) have a good predictiveness in their congeneric groups, also with ϕ_0 is well described hydrophobicity of the bile acids steroid skeleton β side. The mutual association of bile acid salts, especially around the first critical micellar concentration, which is through to their hydrophobic β side of steroid skeleton (Fig. 11) indicated that the chromatographic parameter ϕ_0 adequately be used to predict its critical micellar concentration. Advantages of chromatographic parameters ϕ_0 (6) with respect to the parameters R_{M0} (5) and $\log k_w$ (8) is reflected in the absence of appropriate extrapolation of retention coefficients. Extrapolation error can be large if the linear area of equations (5) and (8) are away from the mobile phase without an organic modifier.

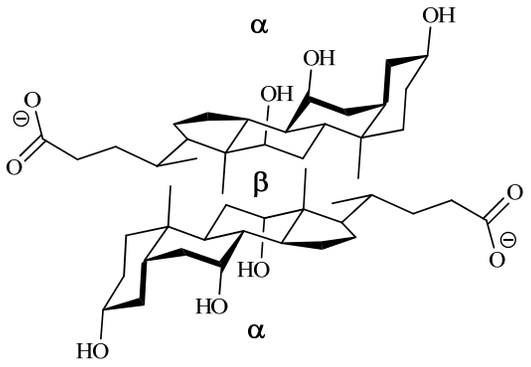


Figure 11. Dimeric micelles of cholic acids anion.

Sarbu et al. in RPTLC experiments examined which chromatographic parameter the best describes the hydrophobicity of bile acids. In the studied bile acids (Table 1) between R_{M0} parameter and the slope S (Equation 5) there is a good correlation, suggesting the presence of congeneric groups among chemical compounds in the tested assembly [3].

Bile acids	Abbreviation	Position and orientation of OHs
Lithocholic acid	**LC**	3α
Deoxycholic acid	**DC**	3α, 12α
Chenodeoxycholic acid	**CDC**	3α, 7α
Ursodeoxycholic acid	**UDC**	3α, 7β
Hyodeoxycholic acid	**HDC**	3α, 6α
Hyocholic acid	**HC**	3α, 6α, 7α
Cholic acid	**C**	3α, 7α, 12α
Glycolithocholic acid	**GLC**	3α
Taurolithocholic acid	**TLC**	3α
Glycochenodeoxycholic acid	**GCDC**	3α, 7α
Taurochenodeoxycholic acid	**TCDC**	3α, 7α
Glycodeoxycholic acid	**GDC**	3α, 12α
Taurodeoxycholic acid	**TDC**	3α, 12α
Glycocholic acid	**GC**	3α, 7α, 12α
Taurocholic acid	**TC**	3α, 7α, 12α

Table 1. Studied bile acids

Application of PCA on the matrix of R_M values (**D**) of the studied bile acids are obtained space of principal components with smaller dimensions then matrix **D**. Screen plot showed that the significant first and second principal components (PC1 and PC2) that explained 99.75% of variance in the starting matrix of retention parameters. Bartlett's statistic shows that PC1, PC2 and PC3 are significant which explained of 99.92% of the total variance. In the set of studied bile acids (Table 1) there is a good correlation between R_{M0} chromatographic parameters and scores of PC1:

$$R_{M0} = 5.379 - 2.624\,\mathrm{PC1} \tag{20}$$

$$\mathrm{R} = -0.9203$$

Compound		R_{M0}	PC1	PC2	PC3	Rodas data:	
			scores			$\log P_{HA}$	$\log P_{HA}$
1	**LC**	5.33	0.195	0.122	0.080		
2	**DC**	4.78	0.346	0.179	0.111	3.50	2.65
3	**CDC**	5.01	0.340	0.177	0.082	3.28	2.25
4	**UDC**	3.65	0.545	0.226	0.114	3.00	2.20
5	**HDC**	3.88	0.488	0.208	0.110	3.08	2.28
6	**HC**	3.66	0.559	0.229	0.107	2.80	1.84
7	**C**	3.84	0.533	0.217	0.105	2.02	1.10
8	**GLC**	3.39	0.674	0.246	0.119		
9	**TLC**	2.12	1.000	0.249	0.077		
10	**GCDC**	2.87	0.819	0.237	0.088	2.12	0.45
11	**TCDC**	2.27	1.249	0.194	0.082		
12	**GDC**	2.90	0.850	0.247	0.087	2.25	0.80
13	**TDC**	2.30	1.165	0.191	0.090		
14	**GC**	2.30	1.102	0.186	0.090	1.65	-0.40
15	**TC**	2.29	1.524	0.137	0.125		

Table 2. Chromatographic parameters

Sarbu for assessing whether R_{M0} or score of PC adequate to describe lipophilicity of bile acids using the experimental data of particion coefficients between 1-octanol and water of bile acids in molecular (HA) and ionized (A) forms. Between the logarithm partition coefficients and chromatographic parameters on the basis of Table 1 are obtained by the following equation of linear regression:

$$\log P_{HA} = 0.366 + 0.620 R_{M0} \tag{21}$$

$$\mathrm{R} = 0.8559$$

$$\log P_{A} = -2.255 + 1.017 R_{M0} \tag{22}$$

$$R = 0.8751$$

$$\log P_{HA} = 3.995 - 2.195\,PC1 \tag{23}$$

$$R = 0.8675$$

$$\log P_{A} = 3.824 - 3.807\,PC1 \tag{24}$$

$$R = 0.9379$$

$$\log P_{HA} = 4.407 - 2.122\,PC1 - 2.168\,PC2 \tag{25}$$

$$R = 0.8715$$

$$\log P_{A} = 3.387 - 3.885\,PC1 + 2.290\,PC2 \tag{26}$$

$$R = 0.9395$$

Linear regression equations (23 and 24) in which the chromatographic parameter R_{M0} replaced with a score of PC1 has correlation coefficients with larger values compared to the equations (21 and 23). Correlation coefficient is slightly increased in the multiple linear regression equation for log*P* (molecular and ionized form) when the equations involved scores of PC1 and PC2. Increasing the coefficient of correlation in the above equations can be explained with Wolds assumption according to which measurement errors are contained in the principal components with small eigenvalues, so the scores of PC1 is purified from the errors. In Fig. 12 bile acids are shown from the Table 2 in the plane of the scores of principal components. It can be seen to form three linear congeneric groups: non-conjugated bile acids (**BA**), glyco-conjugate (**G-BA**) and tauro-conjugated (**T-BA**). Grouping of bile acids on the basis of the scores of PC as the appropriate regression equation for log*P* (from 23 to 26), indicating that the PC1 scores is adequate replacement for R_{M0} chromatographic parameters to describe the lipophilicity (hydrophobicity) of bile acids, namely the scores of PC1 can be used as molecular descriptors for hydrophobicity of the steroid skeleton β (bile acids).

If increase the set of bile acids from Table 1 by including oxo and acetoxy derivatives, R_{M0} chromatographic parameter also included in the multiple linear regressions for the logP_{HA} and the logP_{A} with molecular descriptors such as polarizability and molar specific sum of absolute values of the charges on each atom of the molecule (electronic parameter), molecular volume, the third order of the connectivity index and Wiener index (shape parameter). Based on the loadings of PC can be concluded that R_{M0} carries different information about the structure of the studied bile acids than the above *in silico* molecular descriptors [6].

Effect of temperature on the chromatographic process in reverse phase chromatography can also be applied to obtain chromatographic parameters that have properties of molecular

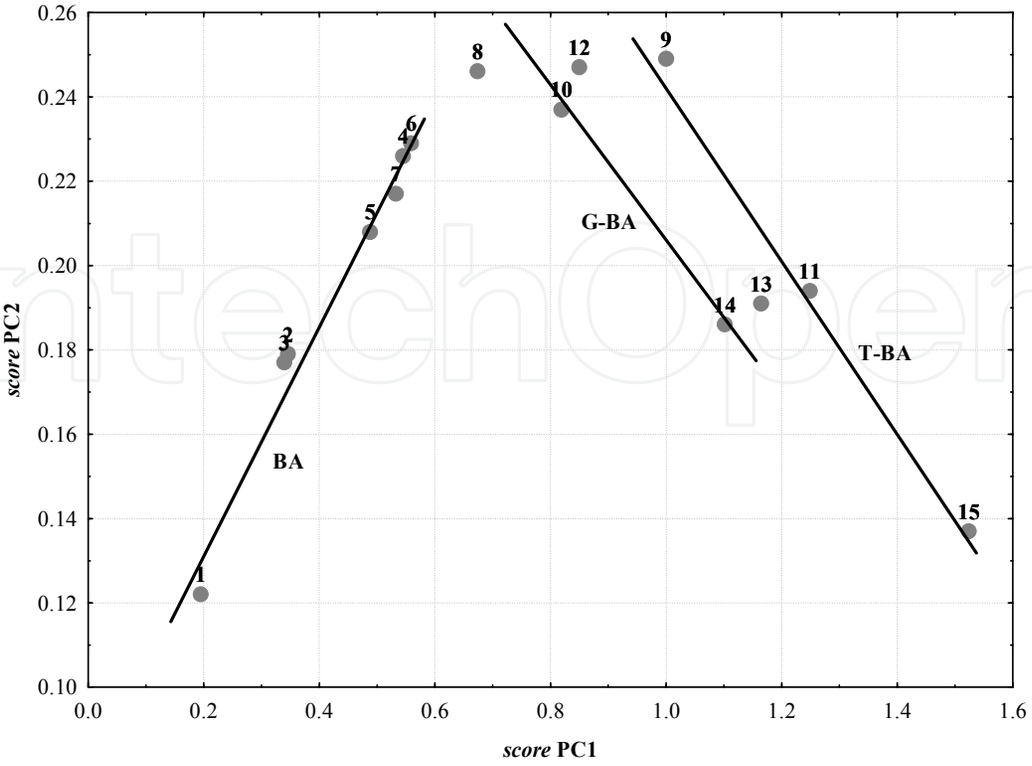


Figure 12. Sore of PC: bile acids from Table 1.

descriptors [15]. Namely, the reaction (R2) is exothermic (H_{ad} <0), then based on the van' t Hoff equation, the equilibrium constant of reaction (R2) decreases

$$\frac{d\ln K_{ad}}{dT} = \frac{dH_{ad}}{RT^2} \tag{27}$$

This means that the retention factor k (9) also decreases. Thus, the increase in temperature has the same effect on the value of k as there are increasing the volume fraction of organic modifier in aqueous mobile phase (reversed phase chromatography). In the experiment, the influence of temperature with an the retention factor at constant mobile phase included 25 different bile acids (conjugated and non-conjugated) (Table 3) [18].

For each tested bile acids between retention factor k and temperature t there is a linear relation:

$$k = a + bt \tag{28}$$

The linear model (28) explains from 96% to 99% of the variance at tested bile acids (Table 4). In the set of the studied bile acids (Table 3) between the parameters of equation (28) (a and b) there is a good correlation, indicating the formation of congeneric groups in the plane of a and b:

$$a = 0.7683 + 58.8522b \tag{29}$$

R=0.9996

Bile acids	Abbreviation	Position and orientation of OH and oxo groups
Lithocholic acid	**LC**	3*α*-OH
Deoxycholic acid	**DC**	3*α*-OH, 12*α*-OH
Chenodeoxycholic acid	**CDC**	3*α*-OH, 7*α*-OH
Cholic acid	**C**	3*α*-OH, 7*α*-OH, 12*α*-OH
Ursodeoxycholic acid	**UDC**	3*α*-OH, 7β-OH
Hyodeoxycholic acid	**HDC**	3*α*-OH, 6*α*-OH
Hyocholic acid	**HC**	3*α*-OH, 6*α*-OH, 7*α*-OH
Glycochenodeoxycholic acid	**GCDC**	3*α*-OH, 7*α*-OH
Taurodeoxycholic acid	**TDC**	3*α*-OH, 12*α*-OH
Glycocholic acid	**GC**	3*α*-OH, 7*α*-OH, 12*α*-OH
Glycodeoxycholic acid	**GDC**	3*α*-OH, 12*α*-OH
Taurolithocholic acid	**TLC**	3*α*-OH
Taurochenodeoxycholic acid	**TCDC**	3*α*-OH, 7*α*-OH
Glycolithocholic acid	**GLC**	3*α*-OH
12-Monoketocholic acid	**12-MKC**	3*α*-OH, 7*α*-OH, 12*α*-Oxo
7-Monoketocholic acid	**7-MKC**	3*α*-OH, 7*α*-Oxo, 12*α*-OH
7,12-Diketocholic acid	**7,12-DKC**	3*α*-OH, 7*α*-Oxo, 12*α*-Oxo
3,7-Diketocholic acid	**3,7-DKC**	3*α*-Oxo, 7*α*-Oxo, 12*α*-OH
3,12-Diketocholic acid	**3,12-DKC**	3*α*-Oxo, 7*α*-OH, 12*α*-Oxo
Dehydrocholic acid	**TKC**	3*α*-Oxo, 7*α*-Oxo, 12*α*-Oxo
12-Monoketodeoxycholic acid	**12-MKD**	3*α*-OH, 12*α*-Oxo
Diketodeoxycholic acid	**DKD**	3*α*-Oxo, 12*α*-Oxo
7-Monoketochenodeoxycholic acid	**7-MKCD**	3*α*-OH, 7*α*-Oxo
Diketochenodeoxycholic acid	**DKCD**	3*α*-Oxo, 7*α*-Oxo
6-Monoketohyodeoxycholic acid	**6-MKHD**	3*α*-OH, 6*α*-Oxo

Table 3. Tested bile acids in the experiment the influence of temperature on the values of retention factor *k*

Fig. 13 shows that the tested bile acids in the plane of parameters of equation (23) formed by three congeneric groups. The first group (**I**) formed by bile acids with oxo groups in the steroid nucleus and the bile acids that in addition to C3 have one more equatorial OH groups. The *α* equatorial OH groups have the same position relative to the steroid skeleton mean plane such have the *α* oxo groups. However, in this group also includes the cholic acid and the glycocholic acid which is structurally not fit into the above mentioned group. The second group (**II**) formed deoxycholic acid and chenodeoxycholic acid with its conjugates. The third group (**III**) forms lithocholic acid and its conjugates. Question may arise as whether can be improved with grouping of bile acid gained by in the plane of *a* and *b*, especially in terms of group **I** which provides the vendor cholic acid? The answer may be the application of methods of principal components of the matrix of the data of retention factor dependence on temperature. The first principal component (PC1) explains 99.99715%

of variance in the starting data matrix, while the remaining variance explained by PC2. In Table 4 presents the scores of principal components of bile acids studied.

Compound		*a*	*b*	R^2	PC1	PC2
					score	
1	**LC**	43.0501	0.7200	0.9879	5.37637	0.023150
2	**DC**	21.8893	0.3518	0.9906	1.79015	-0.051081
3	**CDC**	18.4260	0.3021	0.9807	1.05575	0.020646
4	**C**	8.4873	0.1319	0.9828	-0.66245	0.016292
5	**UDC**	4.4884	0.0680	0.9843	-1.40668	0.031446
6	**HDC**	5.6203	0.0804	0.9538	-1.11512	-0.034642
7	**HC**	5.5477	0.0864	0.9836	-1.23205	0.041543
8	**GCDC**	16.6862	0.2651	0.9946	0.85970	-0.024730
9	**TDC**	19.5503	0.3165	0.9802	1.30724	-0.001736
10	**GC**	7.8569	0.1213	0.9841	-0.77443	0.014749
11	**GDC**	20.6286	0.3336	0.9786	1.51098	-0.012582
12	**TLC**	39.1628	0.6544	0.9728	4.65836	0.029036
13	**TCDC**	16.3808	0.2671	0.9796	0.70668	0.007651
14	**GLC**	41.1016	0.6926	0.9750	4.95053	0.026326
15	**12-MKC**	3.1340	0.0402	0.9704	-1.58139	-0.004996
16	**7-MKC**	2.8202	0.0353	0.9676	-1.64238	-0.002356
17	**7,12-DKC**	0.7141	0.0078	0.9769	-2.12906	0.056420
18	**3,7-DKC**	0.6100	0.0054	0.9974	-2.11715	0.056756
19	**3,12-DKC**	0.6237	0.0060	0.9752	-2.12981	0.056122
20	**TKC**	0.5042	0.0056	0.9738	-2.17086	0.066570
21	**12-MKD**	7.7659	0.1040	0.9652	-0.57871	-0.116428
22	**DKD**	4.1490	0.0511	0.9681	-1.32315	-0.048688
23	**7-MKCD**	6.1746	0.0817	0.9665	-0.91824	-0.078586
24	**DKCD**	3.4944	0.0423	0.9704	-1.46618	-0.031508
25	**6-MKHD**	6.3614	0.0913	0.9667	-0.96809	-0.039373

Table 4. The parameters of the linear model (28) and scores of principal components

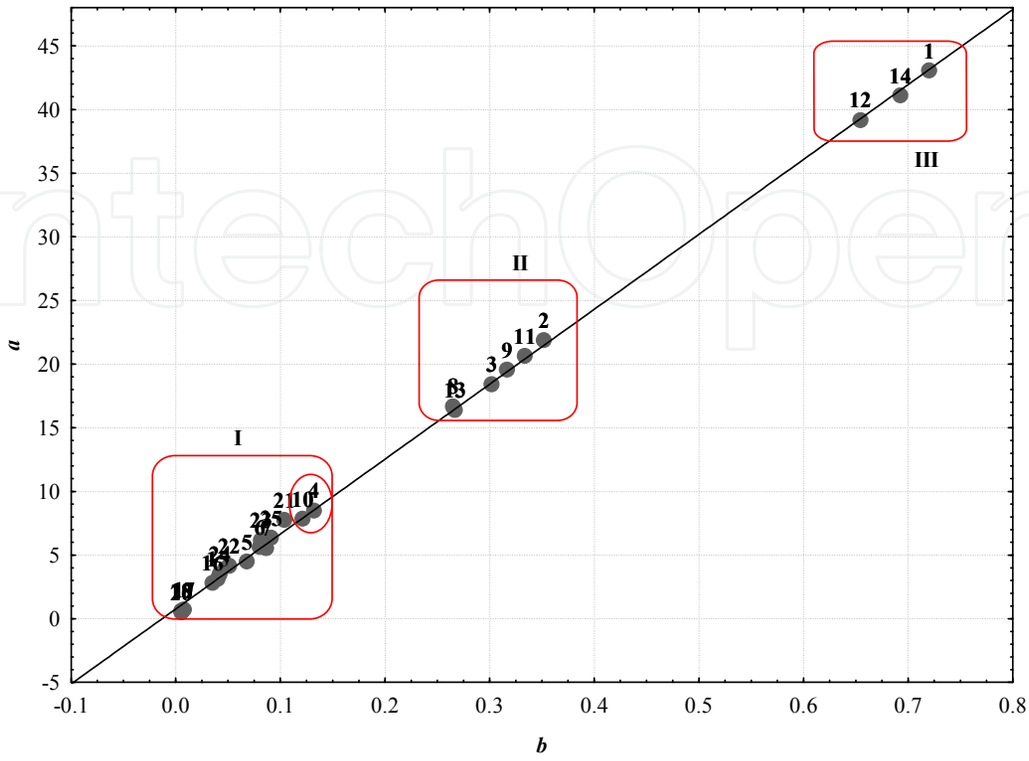


Figure 13. Grouping the bile (Table 4) acid on the basis of linear regression parameters of the retention factor temperatures dependence.

In the plane of principal component scores are obtained by analogous groups such as Fig. 14. However in the group **I** the oxo derivatives and the derivatives with α equatorial OH groups forming linear congeneric group, while based on Cooks distance cholic acid and glycocholic acid may be considered as outliners. With loadings of principal component can be determined by mutual correlation of different *in silico* descriptors (topological and electronic) and obtained experimental descriptors *a*, *b*, and PC1 on the set of 25 bile acids (Table 4). It may be concluded that chromatographic parameters *a* and PC1 are more or less orthogonal to most topological descriptors (Winers index, Balabans index, shape attributes, konectivities, etc.), while PC1 and *a* are mutually parallel so that explain the same structural characteristics of bile acids. Namely, as flat molecules nitrazepan guest molecule is incorporated in the hydrophobic domain of micelles. Therefore, if the β side of bile acids steroid skeleton more hydrophobic then nitrazepans partition coefficient has a higher value. Using the method of partial least square (PLS) regression equation was obtained for the nitrazepans partition coefficients in which experimentally determined variables *a* and PC1 are the most significant with respect to *in silico* descriptors. Which means that PC1 and *a* can be used as molecular descriptors of bile acids lipophilicity, and that the temperature dependence of retention factors may obtain information regarding the structure of bile acids (lipophilicity).

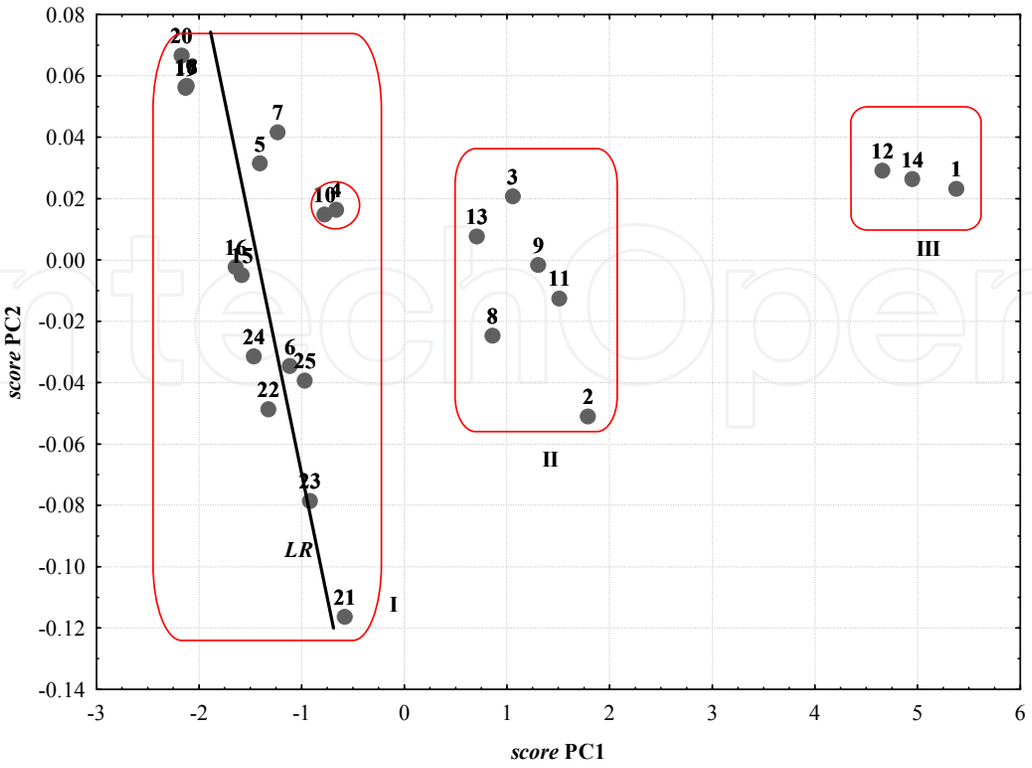


Figure 14. Grouping of bile acids (Table 4) based on the scores of principal component.

6. Conclusions

Chromatographic data (parameters) can be used for prediction of following feaatures of bile acids:

- Critical Micellar Concentration
- Haematolityc potentitial
- Binding for Albumin
- Solubilization of cholesterol, lecitin, vitamin E etc.
- Mixed micelles parameters (for example partition coefficient of nitrazepan in bile acids' micelles)

Application of retention parameters for prediction of bile acids' properties is convenient for finding the size of the starting set of molecules. If the set is too small (less than ten molecules), by using *in silico* descriptors multiple regression equations are obtained which are overfitted, while by using chromatographic parameters that posses appropriate structural informations about bile acids, simple linear equations are derived with acceptable predictive power.

Author details

Mihalj Poša
Department of Pharmacy, Faculty of Medicine, University of Novi Sad, Serbia

Acknowledgement

This study was supported by the Secretariat for Science and Technological Development of AP Vojvodina, Republic of Serbia (Grant No. 114-451-2113/2012).

Abbreviations

QSA(P)R = Quantitative Structure Activity (Property) Relationship
RPHPLC = Reversed Phase High Pressure (Resolutions) Liquid Chromatography
RPTLC = Reversed Phase Thin Layer Chromarography
PCA = Principal Component Analysis
PC = Principal Component
SWM = Stabilised Water Molecules
NSWM = Non Stabilised Water Molecules
SSMP = steroid skeleton mean plane

7. References

[1] Camile W (2003) The Practice of Medicinal Chemistry. Oxford: Academic Press.
[2] King F.D (2002) Medicinal Chemistry Principles and Practice. Cambridge: R.S.C.
[3] Sârbu C, Kuhajda K, Kevrešan S (2001) Evaulation of the lipofilicity of bile acids and their derivatives by thin-layer chromatography and principal component analysis. J. Chromatogr. A 917: 361-366.
[4] Poša M (2011) QSPR study of the effect of steroidal hydroxy and oxo substituents on the critical micellar concentrations of bile acids. Steroids 76(1-2): 85-93.
[5] Poša M, Kuhajda K (2010) Hydrophobiciti and haemolytic potential of oxo derivatives of cholic, deoxycholic and chenodeoxycholic acids. Steroids 75(6): 424-431.
[6] Sârbu C, Onişor C, Poša M, Kuhajda K, Kevrešan S (2008) Modeling and prediction (correction) of partition coefficients of bile acids and their derivatives by multivariate regression methods. Talanta 75: 651-657.
[7] Onişor C, Poša M, Sârbu C (2011) Modeling of chromatographic lipophilicity indices of bile acids and their derivatives. Rev. Roum. Chim. 56(3): 217-229.
[8] Onişor C, Poša M, Kevrešan S, Kuhajda K, Sârbu C (2010) Modeling chromatographic lipophilicity of bile acids and their derivatives estimated by HP-RPTLC. J. Sep. Sci. 33(20): 3110-3118.
[9] Natalini B, Sardella R, Camaioni E, Gioiello A, Pellicciari R (2007) Correlation between CMC and chromatographic index: simple and effective evaluation of the hydrophobic/hydrophilic balance of bile acids. Anal. Bioanal. Chem. 388: 1681-1688.
[10] Natalini B, Sardella R, Camaioni E, Macchiarulo A, Gioiello A, Carbone G, Pellicciari R (2009) Derived chromatographic indices as effective tools to study the self-aggregation process of bile acids. J. Pharm. Biomed. Anal. 50: 613-621.
[11] Poša M, Pilipović A, Lalić M (2010) The influence of NaCl on hydrophobicity of selected, pharmacologically active bile acids expressed with chromatographic retention index and critical micellar concentration. Colloids Surf. B: Biointerfaces 81: 336-343.

[12] Roda A, Minutello A, Angellotti M.A, Fini A (1990) Bile acid structure-activity relationships: evaluation of bile acids lipophilicity using 1-octanol/water partition coefficient and reverse phase HPLC. J. Lipid Res. 31: 1433-1443.
[13] Mikov M, Fawcett J.P (2007) Bile Acids. Geneva: Medishet Publisher.
[14] Poša M, Farkaš Z (2010) Cholesterol solubilization by oxo derivatives of selected bile acids and their osmotic resistance. Collect. Czech. Chem. Commun. 75(8): 767-784.
[15] Poša M, Pilipović A, Lalić M, Popović J (2010) Hydrophobicity and retention coefficient of selected bile acid oxo derivatives. Acta Chim. Slov. 57(4): 828-835.
[16] Heuman M.D (1989) Quantitative estimation of the hydrophilic-hydrophobic balance of mixed bile salt solutions. J. Lipid Res. 30: 719-730.
[17] Armstrong J.M, Carey C.M (1982) The hydrophobic-hydrophilic balance of bile salts. Inverse correlation between reverse-phase high performance liquid chromarographic mobilities and micellar cholesterol-solubilizing capacities. J. Lipid Res. 23: 70-80.
[18] Poša M, Pilipović A, Lalić M, Popović J (2011) Determination and importance of temperature dependence of retention coefficient (RPHPLC) in QSAR model of nitrazepams' partition coefficient in bile acid micelles. Talanta 83(5): 1634-1642.